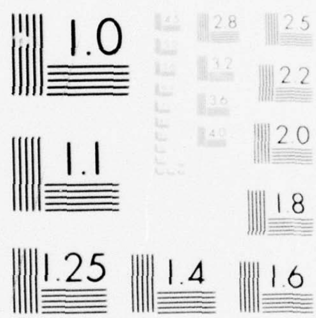
AD-A044 723 MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF MATHEMATICS F/G 12/1
LINEAR FUNCTION OF CONCOMITANTS OF ORDER STATISTICS.(U)
SEP 77 S S YANG N00014-75-C-0555
UNCLASSIFIED TR-7 NL

| OF |
AD
A044723

END
DATE
FILMED
10-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A044723

12

6

LINEAR FUNCTION OF CONCOMITANTS OF ORDER STATISTICS

BY

10 S. S. YANG

14 TR-7

9 TECHNICAL REPORT NO. 7

11 9 SEPTEMBER 9, 1977

12 41p.

PREPARED UNDER CONTRACT

15 N00014-75-C-0555 (NR-042-331)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

220 021

LINEAR FUNCTION OF CONCOMITANTS OF ORDER STATISTICS

by

S. S. Yang

Technical Report No. 7

September 9, 1977

Prepared Under Contract
N00014-75-C-0555 (NR-042-331)
For the Office of Naval Research

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

1. Introduction. Let (X_i, Y_i) $(i=1,2,\ldots,n)$ be i.i.d. as (X,Y). If the X_i are arranged in ascending order as the order statistics $X_{r:n}$ $(r=1,2,\ldots,n)$, the Y-variate associated with $X_{r:n}$ may be denoted by $Y_{[r:n]}$ and termed the concomitant of the r^{th} order statistic. The class of statistics which are linear functions of order statistics has received considerable attention in recent years. A great deal of efforts have been spent on looking into conditions under which such statistics are asymptotically normal. However hardly any work has been done on linear function of the concomitants of order statistics $n^{-1}\sum_{i=1}^{n} C_{in}Y_{[i:n]}$. Recently Bhattacharya (1974,1976) and Sen (1976) have considered asymptotic behavior of the sums $n^{-1}\sum_{i=1}^{[nt]} Y_{[i:n]}$ and $n^{-1}\sum_{F(X_i:n)\le t} Y_{[i:n]}$, where they called the $Y_{[i:n]}$ the "induced order statistics". It is the purpose of this paper to investigate the asymptotic behavior and statistical applications of statistics of the form $n^{-1}\sum_{i=1}^{n} C_{in}Y_{[i:n]}$.

In section 2, we shall consider statistics of the form $S_n = n^{-1}\sum_{i=1}^{n} J(\frac{i}{n+1})Y_{[i:n]}$ and more generally $\phi_n = n^{-1}\sum_{i=1}^{n} J(\frac{i}{n+1})H(X_{i:n}, Y_{[i:n]})$, where J is a smooth bounded function and $H(x,y)$ is some real-valued function

defined on the 2-dimensional Euclidean space. Section 2.1 to 2.2 introduces the paper notation and gives a number of results (some known) which are needed in later proofs. These sections may be skimmed and referred back when necessary. In section 2.3, under faily mild conditions, the asymptotic normality of S_n is established. The asymptotic formulas for the mean and variance of S_n are also obtained. The same asymptotic results are also obtained for ϕ_n. The method of proof adopted is essentially same as that used by Stigler (1969, 1974) for proving asymptotic normality of linear function of order statistics. In fact it is closely in line with his 1974 paper.

In section 3, two applications are also considered. First, the technique developed and results obtained in section 2 are applied to construct consistent estimators of the form $M_n = n^{-1} \sum_{i=1}^{n} J_n(\frac{i}{n}) H(X_{i:n}, Y_{[i:n]})$, where J_n "tends" to a Dirac delta function as $n \to \infty$, for estimating various conditional quantities, for example, $E(Y|X=x)$, $P(Y \varepsilon A|X=x)$ and $Var(Y|X=x)$. Under certain regularity conditions, the asymptotic normality of M_n is also established. Second, based on a statistics of the form ϕ_n a large sample test is proposed for testing a one-sided hypothesis concerning a specified regression function.

2. Asymptotic Normality.

2.1 Notation. Let (X_i, Y_i) $(i=1,2,...,n)$ be i.i.d. as (X,Y).

Then the r^{th} ordered X-variate is denoted by $X_{r:n}$ and the associated Y-variate, the concomitant of the r^{th} ordered statistic, by $Y_{[r:n]}$. For convenience the following notation will also be adopted through out this paper:

$F(x)$ = cdf of the X-variate .

$F_{r:n}(x)$ = cdf of the r^{th} order statistics $X_{r:n}$.

$m(x) = E(Y|X=x)$.

$\sigma^2(x) = Var(Y|X=x)$.

$F^{-1}(u) = \inf\{x|F(x)=u\}$.

$J(u)$ is continuous a.e. $moF^{-1} = J(u)$ and $moF^{-1}(u)$ possess no common discontinuities.

2.2 Some useful results. In this section we shall present several results which will be used in the proofs of the main results of this section.

As in Stigler (1969 and 1974), we shall use Hájek's projection lemma to obtain a sum of independent identically distributed random variables which is asymptotically equivalent in mean squares to S_n . For easier reference, we shall state the lemma as Proposition 1.

Proposition 1. (Hájek 1968). Let $Z_1,\ldots,Z_n$ be independent random variables and $S = S(Z_1,\ldots,Z_n)$ be a statistic depending on $Z_1,\ldots,Z_n$ such that $E\,S^2 < \infty$. Define the projection of S to be

(1) $$\hat{S} = \sum_{i=1}^{n} E(S|Z_i) - (n-1)\, E\,S \ .$$

Then

(2) $$E\,\hat{S} = E\,S$$

and

(3) $$E(S-\hat{S})^2 = \operatorname{Var} S - \operatorname{Var} \hat{S} \ .$$

The following three propositions will be used repeatedly in the derivations of our main results.

Proposition 2. Suppose $h(x)$ can be expressed as a difference of two increasing right continuous functions of the real line, Y is any random variable with cdf G, and $E|h(Y)| < \infty$. Then

(4) $$\int_a^{\infty} h(y)dG(y) = \int_a^{\infty} [1-G(y)]dh(y) + h(a)[1-G(a)]$$

(5) $$\int_{-\infty}^{a} h(y)dG(y) = -\int_{-\infty}^{a} G(y)dh(y) + h(a)G(a) \ .$$

Proof. We shall only consider (4). Similar arguments can be used to show (5). Without loss of generality, we can assume $h(y)$ is an increasing right continuous function of y. Then we may write

$$\int_a^{\infty} [h(y)-h(a)]dG(y)$$

$$= \int_a^{\infty} \left\{ \int_{-\infty}^{\infty} [I_{(t \le y)} - I_{(t < a)}]dh(t) \right\} dG(y) .$$

Since $E|h(Y)| < \infty$, by Fubini's Theorem we can interchange order of integrations and the resulting expression is reduced to

$$\int_a^{\infty} [1-G(t)]dh(t) .$$

Clearly the resulting equation yields (4).

Proposition 3. Let X and Y be any two random variables with joing distribution given by $G(x,y)$ and marginal distributions $F(x)$ and $H(x)$ respectively, and $g(x)$ be any real-valued function which can be expressed as a difference of two increasing right continuous functions. Then

$$(6) \qquad \mathrm{Cov}(g(X),g(Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x,y)-F(x)H(y)]dg(x)dg(y)$$

provided $E(g(X)g(Y))$, $E(g(X))$ and $E(g(Y))$ exist.

The proof is essentially the same as that given in Lehman (1966) for proving the special case $g(x) = x$, which is attributed to Hoeffding (1948).

Proposition 4. Assume $J(u)$ is bounded and continuous a.e. on F^{-1}, $g(x)$ is a real valued function and X is a random variable with distribution $F(x)$ such that $E|g(X)| < \infty$. Then

$$(7) \qquad \lim_{n\to\infty} E\left[n^{-1}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) g(X_{i:n})\right]$$

$$= \int_{-\infty}^{\infty} J(F(x))g(x)\,dF(x) \ ,$$

provided that one of the following conditions holds:

(a) $F(x)$ is absolutely continuous.

(b) $g(x)$ can be expressed as a difference of two increasing right continuous functions.

Proof. Suppose (a) holds, then

$$E\left[n^{-1}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) g(X_{i:n})\right]$$

$$= \int_{-\infty}^{\infty} g(x)\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)\binom{n-1}{i-1} F(x)^{i-1}[1-F(x)]^{n-i}\,dF(x) \ .$$

Hence by Bernoulli's weak law of large number and Dominated Convergence Theorem, we can immediately establish (7).

Now suppose (b) holds. By Proposition 2 and the fact that

$F_{i:n}(t) = \int_0^{F(t)} n\binom{n-1}{i-1}u^{i-1}(1-u)^{n-i}\,du$, we may write

$$(8) \qquad E\left[n^{-1}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)g(X_{i:n})\right]$$

$$= \int_0^{\infty}\left[\int_t^{\infty} Q_n(F(x))\,dF(x)\right]dg(t) + g(0)\int_0^{\infty} Q_n(F(x))\,dF(x)$$

$$- \int_{-\infty}^{0}\left[\int_{-\infty}^{t} Q_n(F(x))\,dF(x)\right]dg(t) + g(0)\int_{-\infty}^{0} Q_n(F(x))\,dF(x) \;,$$

where $Q_n(u) = \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)\binom{n-1}{i-1}u^{i-1}(1-u)^{n-i}$.

By changing order of integration, the right hand side of (8) can be simplified to

$$\int_{-\infty}^{\infty} g(x)Q_n(F(x))\,dF(x) \;.$$

Hence (7) holds.

The following representations for the moments of the concomitants $Y_{[r:n]}$ $(r=1,2,\ldots,n)$ are needed for obtaining a simple expression for $\mathrm{Var}(S_n)$.

<u>Proposition 5</u>. Let $F_{r_1,\ldots,r_{k:n}}(x_1,\ldots,x_k)$ be the

joint cdf of the k ordered statistics $X_{r_1:n} \leq X_{r_2:n} \leq \cdots \leq X_{r_k:n}$. Then

(10) $$P(Y_{[r_1:n]} \leq y_1, \ldots, Y_{[r_k:n]} \leq y_k)$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{k} P(Y_i \leq y_i | X_i = x_i) dF_{r_1,\ldots,r_k:n}(x_1,\ldots,x_k) ,$$

and furthermore,

(11)
$$E(Y_{[r:n]}) = E(m(X_{r:n}))$$

$$Var(Y_{[r:n]}) = Var(m(X_{r:n})) + E(\sigma^2(X_{r:n}))$$

$$Cov(Y_{[r:n]}, Y_{[s:n]}) = Cov(m(X_{r:n}), m(X_{s:n})) \quad (r \neq s) .$$

Proof. Since (X_i, Y_i) $(i=1,2,\ldots,n)$ are independent and identically distributed,

$$P(Y_{[r:n]} \leq y | X_{r:n} = x , \quad X_{r:n} = X_i)$$

$$= P(Y_1 \leq y | X_1 = x) ,$$

which does not depend on i . This implies that

$$P(Y_{[r:n]} \leq y \mid X_{r:n}=x) = P(Y \leq y \mid X = x)$$

and $P(Y_{[r:n]} \leq y) = \int_{-\infty}^{\infty} P(Y \leq y \mid X = x)\, dF_{r:n}(x)$.

Similarly we can show (10) and thus (11) can also be easily established.

In order to describe the projection $\hat{Y}_{[r:n]}$ of the $Y_{[r:n]}$ and their covariances, we need the following results.

Proposition 6. Suppose $E(Y^2_{[r:n]} + Y^2_{[s:n]}) < \infty$ for $1 < r$, $s < n$, and $m(x)$ can be expressed as a difference of two increasing right continuous functions. Then for $1 \leq i \leq n$,

$$\begin{aligned} (12)\qquad & E(Y_{[s:n]} \mid X_i=x_i\ ,\ \ Y_i=y_i) \\ & = (y_i-m(x_i))P_s(x_i) + \int_{-\infty}^{\infty} I(t \leq x_i) P_s(t)\, dm(t) \\ & + E(Y_{[s-1:n-1]})\ , \end{aligned}$$

and

$$\begin{aligned} & \mathrm{Cov}[E(Y_{[r:n]} \mid X_i, Y_i),\ E(Y_{[s:n]} \mid X_i, Y_i)] \\ (13)\qquad & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)] P_r(x) P_s(y)\, dm(x)\, dm(y) \\ & + \int_{-\infty}^{\infty} P_r(x) P_s(x) \sigma^2(x)\, dF(x)\ , \end{aligned}$$

where $P_s(x) = \binom{n-1}{s-1} F(x)^{s-1}[1-F(x)]^{n-s}$.

Proof. For any fixed n , let $X_{1:n-1} \le \cdots \le X_{n-1:n-1}$ be the order statistics of $X_1, X_2, \ldots, X_{n-1}$; and let $Y_{[1:n-1]}, \ldots, Y_{[n-1:n-1]}$ be the corresponding concomitants of order statistics. Now for $1 < s < n$,

$$
\begin{aligned}
& E(Y_{[s:n]} | X_n = x_n,\ Y_n = y_n) \\
= \ & E[Y_{[s:n]} I(X_{s-1:n-1} \le X_n \le X_{s:n-1}) | X_n = x_n,\ Y_n = y_n] \\
+ \ & E[Y_{[s:n]} I(X_n < X_{s-1:n-1}) | X_n = x_n,\ Y_n = y_n] \\
+ \ & E[Y_{[s:n]} I(X_n > X_{s:n-1}) | X_n = x_n,\ Y_n = y_n] \\
= \ & y_n P_s(x_n) + E(Y_{[s-n:n-1]}) - \int_{-\infty}^{x_n} m(x)\, dF_{s-1:n-1}(x) \\
+ \ & \int_{-\infty}^{x_n} m(x)\, dF_{s:n-1}(x) \ .
\end{aligned}
$$

Applying Proposition 2 to the last two integrals, the resulting expression can be simplified to

$$
\begin{aligned}
& E(Y_{[s-1:n-1]}) + (y_n - m(x_n)) P_s(x_n) \\
+ & \int_{-\infty}^{\infty} I(t \le x_n) P_s(t)\, dm(t) \ .
\end{aligned}
$$

A symmetric argument gives (12). Now using conditional argument, conditioning on X_i, we can easily show that

$$\mathrm{Cov}[E(Y_{[r:n]}|X_i,Y_i),\ E(Y_{[s:n]}|X_i,Y_i)]$$

$$= E[P_r(X_i)P_s(X_i)\sigma^2(X_i)]$$

$$+ \mathrm{Cov}\Big[\int_{-\infty}^{\infty} I(t \le X_i)P_r(t)dm(t)\ ,\ \int_{-\infty}^{\infty} I(t \le X_i)P_s(t)dm(t)\Big]\ .$$

By Fubini's theorem, the covariance term can be written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]P_r(x)P_s(y)dm(x)dm(y)\ .$$

This completes the proof.

We shall see later that it is more convenient to consider

$T_n = n^{-1} \sum_{i=2}^{n-1} J(\frac{i}{n+1})\ Y_{[i:n]}$ than $S_n = n^{-1} \sum_{i=1}^{n} J(\frac{i}{n+1})\ Y_{[i:n]}$.

The following proposition implies that the contribution of $Y_{[1:n]}$ and $Y_{[n:n]}$ to S_n is asymptotically negligible.

<u>Proposition 7</u>. Suppose $E(Y^2) < \infty$, and either $F(x)$ is absolutely continuous or $E(Y^2|X=x)$ can be expressed as a

difference of two increasing right continuous functions. Then $n^{-1} E(Y^2_{[1:n]} + Y^2_{[n:n]}) \to 0$ as $n \to \infty$.

Proof. Let $g(x) = E(Y^2|X=x)$. Suppose $F(x)$ is absolutely continuous. Then we may write

$$n^{-1} E(Y^2_{[n:n]}) = \int_{-\infty}^{\infty} g(x)F(x)^{n-1}dF(x) ,$$

and clearly by Dominated Convergence Theorem it tends to zero as $n \to \infty$. Now if the condition of $E(Y^2|X=x)$ is satisfied, then by Proposition 2 we may write

$$n^{-1}E(Y^2_{[n:n]})$$

$$= \int_0^{\infty} n^{-1}[1-F_{n:n}(x)]dg(x) - \int_{-\infty}^{0} n^{-1}F_{n:n}(x)dg(x) + n^{-1}g(0) .$$

Clearly the above two integrands tend to zero for each x as $n \to \infty$. Since $n^{-1}[1-F_{n:n}(x)] \leq [1-F(x)]$ and $n^{-1}F_{n:n}(x) \leq F(x)$ for all x , and $\int_0^{\infty}[1-F(x)]dg(x)$ and $\int_{-\infty}^{0} F(x)dg(x)$ are finite by hypothesis, by Dominated Convergence Theorem $n^{-1}E(Y^2_{[n:n]}) \to 0$ as $n \to \infty$. A symmetric argument yields $n^{-1}E(Y^2_{[1:n]}) \to 0$ as $n \to \infty$.

2.3 Main result. Now we are ready to prove the main result of section 2. We shall show that under certain regularity conditions, S_n is asymptotically normally distributed. However we may consider $T_n = n^{-1} \sum_{i=2}^{n-1} J(\frac{i}{n+1}) Y_{[i:n]}$ rather than S_n because T_n and S_n, by Proposition 7, are asymptotically equivalent in mean square. Theorem 1 below and Proposition 1 will show that T_n and $\hat{T}_n$ are asymptotically equivalent. Theorem 2 will then establish the asymptotic normality of $\hat{T}_n$ and thus of S_n. In general, it turns out that under certain regularity conditions, exactly the same method of proof can establish the asymptotic normality of statistics of the form $\phi_n = n^{-1} \sum_{i=1}^{n} J(\frac{i}{n+1}) H(X_{i:n}, Y_{[i:n]})$ where $H(x,y)$ is some real valued function defined on the 2-dimensional Euclidean space.

Theorem 1. Suppose $E|Y^2| < \infty$, $J(u)$ is bounded and continuous a.e. $m \circ F^{-1}$, $m(x)$ can be expressed as a difference of two increasing right continuous function, and either $F(x)$ is absolutely continuous or $\sigma^2(x)$ has the same property of $m(x)$. Then

(14) $$\lim_{n\to\infty} n \operatorname{Var}(\hat{T}_n) = \sigma^2(J,F) ,$$

(15) $$\lim_{n\to\infty} n \operatorname{Var}(T_n) = \sigma^2(J,F) ,$$

where $\hat{T}_n$ is defined as in Proposition 1, and

$$(16) \qquad \sigma^2(J,F) = \int_{-\infty}^{\infty} J(F(x))^2 \, \sigma^2(x)\,dF(x) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]J(F(x))J(F(y))\,dm(x)\,dm(y) \; .$$

Proof. We shall first show (14). Let $\hat{Y}_{[r:n]}$ be defined as in Proposition 1. Then by Proposition 6, we have

$$\mathrm{Cov}(\hat{Y}_{[r:n]}, \hat{Y}_{[s:n]}) = n\left\{ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]P_r(x)P_s(y)\,dm(x)\,dm(y) + \int_{-\infty}^{\infty} P_r(x)P_s(x)\,\sigma^2(x)\,dF(x)\right\}$$

Then

$$n\sigma^2(\hat{T}_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]Q_n(x)Q_n(y)\,dm(x)\,dm(y) + \int_{-\infty}^{\infty} Q_n(x)^2 \, \sigma^2(x)\,dF(x)$$

where $Q_n(x) = \sum_{i=2}^{n-1} J(\frac{i}{n+1})P_i(x)$.

Now by the Bernoulli's weak law of large numbers, $Q_n(x) \to J(F(x))$ for every x such that $J(u)$ is continuous at $u = F(x)$ and $0 < F(x) < 1$.

Since $|Q_n(x)| \leq \sup_x |J(x)|$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)]\,dm(x)\,dm(y) + \int_{-\infty}^{\infty} \sigma^2(x)\,dF(x)$$

$$= \mathrm{Var}(Y)$$

is assumed finite, it follows from the Dominated Convergence Theorem that $n\,\mathrm{Var}(\hat{T}_n) \to \sigma^2(J,F)$ as $n \to \infty$.

Now we shall prove (15). Since by Proposition 7 $n^{1/2}T_n$ and $n^{1/2}S_n$ are asymptotically equivalent in mean square, it suffices to show that $n\,\mathrm{Var}(S_n) \to \sigma^2(J,F)$ as $n \to \infty$.

By equations (11) of proposition 5, we may write

$$(17) \qquad n\,\mathrm{Var}(S_n) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} J\left(\frac{i}{n+1}\right) J\left(\frac{j}{n+1}\right) \mathrm{Cov}[m(X_{i:n}), m(X_{j:n})]$$

$$+ n^{-1} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)^2 E[\mathrm{Var}(Y|X = X_{i:n})] .$$

By Proposition 4 the second sum on the right hand side of (17) converges to

$$\int_{-\infty}^{\infty} J(F(x))^2 \sigma^2(x)\,dF(x) \quad \text{as} \quad n \to \infty .$$

Let

$$G_i(x) = P(X_{i:n} \leq x) \quad ,$$

$$G_{ij}(x,y) = P(X_{i:n} \leq x \; , \; X_{j:n} \leq y) \quad .$$

Then by Proposition 3

$$\mathrm{Cov}[m(X_{i:n}), m(X_{j:n})]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [G_{ij}(x,y) - G_i(x)G_j(y)]dm(x)dm(y) \quad ,$$

and hence the first summation term on the right hand side of (17) can be written as

(18) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_n(x,y)dm(x)dm(y)$$

where $H_n(x,y) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} J(\frac{i}{n+1})J(\frac{j}{n+1})[G_{ij}(x,y) - G_i(x)G_j(y)]$.

Hence to prove (15) it remains to show that

(19) $$\lim_{n\to\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_n(x,y)dm(x)dm(y)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)]J(F(x))J(F(y))dm(x)dm(y) \quad .$$

The proof is identical to that of equation (9) in Theorem 1 of Stigler (1974) and is omitted here. This completes the proof of the Theorem.

<u>Theorem 2</u>. $\mathcal{L}\left[\frac{(S_n - E(S_n))}{\sqrt{\mathrm{Var}(S_n)}}\right] \longrightarrow N(0,1)$ as $n \to \infty$,

provided that the following conditions are satisfied:

(i) $\sigma^2(J,F) > 0$, where $\sigma^2(J,F)$ is given in (16).

(ii) $E(Y^2) < \infty$.

(iii) $J(u)$ is bounded and continuous a.e. $m o F^{-1}$.

(iv) $m(x)$ can be expressed as difference of two increasing right continuous functions.

(v) Either $\sigma^2(x)$ has the same property of $m(x)$ or $F(x)$ is absolutely continuous.

<u>Proof</u>: Propsosition 7 implies that $n^{1/2}(S_n - E(S_n))$ and $n^{1/2}(T_n - E(T_n))$ are asymptotically equivalent in mean square, and Theorem 1 and Proposition 1 imply that of $n^{1/2}(T_n - E(T_n))$ and $n^{1/2}(\hat{T}_n - E(\hat{T}_n))$. Hence it only remains to show that $n^{1/2}(\hat{T}_n - E(\hat{T}_n))$ is asymptotically normally distributed. By Proposition 6 we may write

$$\hat{T}_n - E(\hat{T}_n) = \frac{1}{n} \sum_{i=1}^{n} (y_i - m(X_i)) Q_n(X_i)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} [F(t) - I(X_i \leq t)] Q_n)(t) \, dm(t)$$

where $Q_n(x) = \sum_{j=2}^{n-2} J(\frac{j}{n+1}) P_j(x)$.

Let $W_n = n^{-1} \sum_{i=1}^{n} Z_i$ where

$$Z_i = [Y_i - m(X_i)] J(F(X_i)) + \int_{-\infty}^{\infty} [F(t) - I(X_i \leq t)] J(F(t)) \, dm(t) .$$

Now ,

$$E[n(\hat{T}_n - E(\hat{T}_n) - W_n)^2] = E\{[Y - m(X)]^2 [Q_n(X) - J(F(X))]^2\}$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)] [Q_n(x) - J(F(x))] [Q_n(y) - J(F(y))]$$

$$dm(x) \, dm(y) .$$

Since $Q_n(t) \to J(F(t))$ a.e. $m o F^{-1}$ as $n \to \infty$, by Dominated Convergence Theorem we can show that $E[n^{1/2}(\hat{T}_n - E(\hat{T}_n) - W_n)]^2 \to 0$ as $n \to \infty$.

Since the Z_i are i.i.d. random variables with zero mean and variance $\sigma^2(J,F)$, by Central Limit Theorem, $n^{1/2} W_n$ is

asymptotically $N(0,\sigma^2(J,F))$, and so is $n^{1/2}(\hat{T}_n - E(\hat{T}_n))$. This completes the proof.

Since statistics of the form S_n are usually used as estimates of their expectation, the limiting value of $E(S_n)$ is of interest. The following theorem, which is a direct consequence of Proposition 4, gives the limiting value of $E(S_n)$.

Theorem 3. Suppose $E(|Y|) < \infty$ and $J(u)$ is bounded and continuous a.e. F^{-1} . Then

$$\lim_{n\to\infty} E(S_n) = \int_{-\infty}^{\infty} J(F(x))m(x)dF(x) \tag{20}$$

provided one of the following conditions is satisfied:

(a) $m(x)$ can be epxressed as a difference of two increasing right continuous function of x .

(b) $F(x)$ is absolutely continuous.

Inspection of the proofs of Theorem 1 through 3 reveals that exactly the same arguments used above can be employed to prove the following theorem concerning the asymptotic distribution of ϕ_n .

Theorem 4. Let $\phi_n = n^{-1} \sum_{i=1}^{n} J(\frac{i}{n+1})H(X_{i:n},Y_{[i:n]})$

$\alpha(x) = E[H(X,Y)|X=x]$ and $\tau^2(x) = E\{[H(X,Y)-\alpha(X)]^2|X=x\}$.

Suppose the following conditions are satisfied:

(i) $J(u)$ is bounded and continuous a.e. moF^{-1} .

(ii) $E(H(X,Y)^2) < \infty$.

(iii) $\alpha(x)$ can be expressed as a difference of two increasing right continuous functions.

(iv) Either $\tau^2(x)$ has the same property of $\alpha(x)$ or $F(x)$ is absolutely continuous.

Then

$$(21) \qquad \lim_{n\to\infty} n\,\mathrm{Var}(\phi_n) = \int_{-\infty}^{\infty} J(F(x))^2\,\tau^2(x)\,dF(x) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]J(F(x))J(F(y))\,d\alpha(x)\,d\alpha(y) ,$$

and

$$(22) \qquad \lim_{n\to\infty} E(\phi_n) = \int_{-\infty}^{\infty} \alpha(x)J(F(x))\,dF(x) .$$

Furthermore if the asymptotic variance is positive, then

$$(23) \qquad \mathcal{L}\left[\frac{(\phi_n - E(\phi_n))}{\sqrt{\mathrm{Var}(\phi_n)}}\right] \longrightarrow N(0,1) \quad \text{as} \quad n \to \infty .$$

<u>Remark 1</u>. Let $H(x,y) = h(x)$. Theorem 4 establishes the asymptotic normality of $A_n = n^{-1} \sum_{i=1}^{n} J(\frac{i}{n+1})h(X_{i:n})$ with asymptotic variance

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x\wedge y)-F(x)F(y)]J(F(x))J(F(y))dh(x)dh(y) \quad .$$

Remark 2. Let $H(x,y) = (y-m(x))$. Then Theorem 4 establishes the asymptotic normality of

$$R_n = n^{-1} \sum_{i=1}^{n} J(\frac{i}{n+1}) \; [Y_{[i:n]}-m(X_{i:n})]$$

with asymptotic variance $\int_{-\infty}^{\infty} J(F(x))^2 \; \sigma^2(x)dF(x)$. Note also that because of the special form of $H(x,y)$, condition (iii) of Theorem 4 can be omitted here.

Remark 3. Let $S_n' = n^{-1} \sum_{i=1}^{n} J_n(\frac{i}{n+1})Y_{[i:n]}$. Examination the proofs of section 2 immediately reveals that Theorem 1 through 4 apply equally well to S_n' as long as J_n is uniformly bounded, and for every continuity point u_0 of J there is an open neighborhood of u_0 in which $J_n(u) \to J(u)$ uniformly.

3. Application

3.1 Estimation. In this section the technique developed and results obtained in section 2 are applied to construct consistent estimators of various conditional quantities based on statistics of the form

$$M_n = n^{-1} \sum_{i=1}^{n} J_n(\frac{i}{n}) \; H(X_{i:n}, Y_{[i:n]}) \quad .$$

Under certain regularity conditions, M_n is also asymptotically normally distributed. We shall first prove the following theorem which will give sufficient condition for M_n to be a consistent estimator for $\alpha(x) = E[H(X,Y)|X=x]$.

Theorem 5. Suppose conditions (ii) to (iv) of Theorem 4 are satisfied. Let $J_n(u) = h(n)^{-1}K\left(\frac{u-F_n(x_0)}{h(n)}\right)$ where $h(n) \to 0$ as $n \to \infty$ and $F_n(x)$ is the emprical distribution of the X_i . In addition, the following conditions are also assumed to be satisfied:

(i) There exists constant $M\ (>0)$ such that $|K(t_1)-K(t_2)| \le M|t_1-t_2|$ for all t_1 and t_2 .

(ii) $|tK(t)| \to 0$ as $|t| \to \infty$.

(iii) $\int_{-\infty}^{\infty} K(t)\,dt = 1$.

(iv) $\lim_{n\to\infty} h(n)^{-1}\left(\frac{\log\log n}{n}\right)^{1/4} = 0$.

(v) $\alpha o F^{-1}(u) = E(H(X,Y)|X=F^{-1}(u))$ is continuous at $u_0 = F(x_0)$.

Then

$$\lim_{n\to\infty} E(M_n) = \alpha(x_0) \,, \tag{24}$$

and

$$\lim_{n\to\infty} E(M_n-\alpha(x_0))^2 = 0 \,. \tag{25}$$

Before proving Theorem 5, we shall first prove the following lemma which will be useful for proving (24) and the next theorem.

Lemma 1. Let $K(t)$ be defined as in Theorem 5 and $p(i;n,u) = \binom{n-1}{i-1} u^{i-1}(1-u)^{n-i}$, $0 \leq u \leq 1$. If condition (iv) of Theorem 5 is satisifed, then

$$\sum_{i=1}^{n} h(n)^{-1}\left[K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) - K\left(\frac{u-u_0}{h(n)}\right)\right] p(i;n,u) \to 0$$

uniformly in u as $n \to \infty$.

Proof. By (i) of Theorem 5, we may write

$$\left|\sum_{i=1}^{n} h(n)^{-1}\left[K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) - K\left(\frac{u-u_0}{h(n)}\right)\right] p(i;n,u)\right|$$

$$\leq M \sum_{i=1}^{n} h(n)^{-2} \left|\frac{i}{n} - u\right| p(i;n,u)$$

Now the Law of Iterated Logarithm for the sum of i.i.d. Bernoulli's random variables and condition (iv) imply the conclusion of the lemma.

Proof of Theorem 5. Let

$$M_n^* = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F(x_0)}{h(n)}\right) H(X_{i:n}, Y_{[i:n]}) .$$

Then M_n^* and M_n are asymptotically equivalent in mean square. To see this, write

$$(26) \qquad h(n)^{-1} \left| K\left(\frac{\frac{i}{n} - F_n(x_0)}{h(n)}\right) - K\left(\frac{\frac{i}{n} - F(x_0)}{h(n)}\right) \right|$$

$$\leq M\, h(n)^{-2} \, |F_n(x_0) - F(x_0)| .$$

Then condition (iv) and the Law of Iterated Logarithm implies that with probability one (26) converges to zero uniformly in i as $n \to \infty$. Now it is easy to check that $E(M_n^* - M_n)^2 \to 0$ as $n \to \infty$.

Hence we may consider M_n^* rather than M_n . We shall first show that $E(M_n^*) \to \alpha(x_0)$ as $n \to \infty$. As in Proposition 4, we may write

$$(27) \qquad E(M_n^*) = \int_{-\infty}^{\infty} Q_n^*(F(x))\alpha(x)\,dF(x) ,$$

where $Q_n^*(u) = \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F(x_0)}{h(n)}\right) \binom{n-1}{i-1} u^{i-1}(1-u)^{n-i}$.

Lemma 1 implies that the limiting value of (27) is same as that of

$$(28) \qquad \int_0^1 h(n)^{-1} K\left(\frac{u - F(x_0)}{h(n)}\right) \alpha \circ F^{-1}(u)\,du .$$

With the given conditions on $K(t)$ and $\alpha o F^{-1}(u)$, it is easy to show that the limiting value of (28) is $\alpha(x_0)$. A proof may essentially be found in, e.g., Parzen (1962).

To prove (25), it remains to show that $\mathrm{Var}(M_n^*) \to 0$ as $n \to \infty$. But by Remark 3, $nh(n)^2 \mathrm{Var}(M_n^*) \to 0$ as $n \to \infty$ which implies that of $\mathrm{Var}(M_n^*)$. This completes the proof of the Theorem.

Remark 4. Now it is clear that if the appropriate conditions are satisified,

$$M_{n1} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F_n(x_0)}{h(n)}\right) Y_{[i:n]} ,$$

$$M_{n2} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F_n(x_0)}{h(n)}\right) I(Y_{[i:n]} \varepsilon A) ,$$

and

$$M_{n3} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F_n(x_0)}{h(n)}\right) Y^2_{[i:n]} - M^2_{n1}$$

are respectively consistent estimators of $m(x_0)$, $P(Y \varepsilon A | X=x_0)$ and $\mathrm{Var}(Y|X=x_0)$.

Next we shall prove the following theorem which will give sufficient conditions for $M_n^* = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - F(x_0)}{h(n)}\right) H(X_{i:n}, Y_{[i:n]})$ to be asymptotically normally distributed. Then the asymptotic normality of M_n will be established by showing that $(h(n)n)^{1/2} M_n$ and $(h(n)n)^{1/2} M_n^*$ are asymptotically equivalent in mean square.

Theorem 6. Let $M_n^* = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) H(X_{i:n}, Y_{[i:n]})$. Suppose the assumptions of Theorem 5 are satisfied, $E[H(X,Y)^3] < \infty$, $\alpha'(x)$ exists and $\alpha' oF^{-1}(u)$ is continuous at u_0 with $0 < u_0 < 1$. Then

$$\mathcal{L}[(n\,h(n))^{1/2} (M_n^* - E(M_n^*))] \to N(0, \gamma^2(u_0)) \quad \text{as} \quad n \to \infty$$

where $\gamma^2(u_0) = \tau^2 oF^{-1}(u_0) \int_{-\infty}^{\infty} K(t)^2 dt$.

Proof. For easier presentation, we shall prove the special case

$$M_n^* = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) Y_{[i:n]} .$$

Similar arguments can be used to prove the general case. Let

$$A_n^* = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) m(X_{i:n}) ,$$

$$B_n^* = M_n^* - A_n^* .$$

Then clearly,

$$(n\,h(n))^{1/2} (M_n^* - E(M_n^*)) = (n\,h(h))^{1/2} (A_n^* - E(A_n^*)) + (n\,h(n))^{1/2} B_n^* .$$

We shall first show that $n\,h(n) \operatorname{Var}(A_n^*) \to 0$ as $n \to \infty$, and

then the asymptotic normality of $(n\,h(n))^{1/2} B_n^*$, which will imply that of $(n\,h(n))^{1/2} (M_n^* - E(M_n^*))$.

As in Theorem 1, we may write

$$n\,h(n)\,\mathrm{Var}(A_n^*) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_n^*(x,y)\,dm(x)\,dm(y) \quad ,$$

where

$$H_n^*(x,y) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) K\left(\frac{\frac{j}{n} - u_0}{h(n)}\right) [G_{ij}(x,y) - G_i(x) G_j(y)] \; ,$$

and $G_{ij}(x,y)$ and $G_i(x)$ are defined as in Theorem 1. Applying the same type of argument used to prove Lemma 1, we can easily show that

$$\left\{ H_n^*(x,y) - h(n)^{-1} K\left(\frac{F(x) - u_0}{h(n)}\right) K\left(\frac{F(y) - u_0}{h(n)}\right) [F(x,y) - F(x)F(y)] \right\}$$

$\to 0$ uniformly in x and y as $n \to \infty$. Hence we have, as $n \to \infty$,

$$(29) \quad n\,h(n)\,\mathrm{Var}(A_n^*)$$

$$= \int_0^1 \int_0^1 h(n)^{-1} K\left(\frac{u - u_0}{h(n)}\right) K\left(\frac{v - u_0}{h(n)}\right) (u \wedge v - uv)\, m' \circ F^{-1}(u)\, m' \circ F^{-1}(v)\, du\, dv$$

$$+ \; o(1) \; .$$

But we can easily show that the limiting value of $h(n)^{-1}$ times

the integral term of the right hand side of (29) is

$$u_0(1-u_0)[m'\circ F^{-1}(u_0)\int_{-\infty}^{\infty} K(t)dt]^2 \quad \text{(cf. Cacoullas, 1966).}$$

Hence $\lim_{n\to\infty} n\,h(n)\,\mathrm{Var}(A_n^*) = 0$.

Let $\hat{B}_n^*$ be the projection of B_n^* . As in Proposition 6, it can be written as

$$\hat{B}_n^* = n^{-1} \sum_{i=1}^{n} (Y_i - m(X_i))Q_n^*(F(X_i)) \ ,$$

where $Q_n^*(t)$ is def ned as in (27) of Theorem 5. Moreover, we can also write

$$\text{(30)} \quad n\,h(n)\,\mathrm{Var}(B_n^*) = \int_0^1 \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right)^2 p(i;n,u)\,\sigma^2\circ F^{-1}(u)du \ ,$$

$$\text{(31)} \quad n\,h(n)\,\mathrm{Var}(\hat{B}_n^*) = \int_0^1 h(n)Q_n^*(u)^2\,\sigma^2\circ F^{-1}(u)du \ .$$

Using similar arguments as that for deriving (24) of Theorem 5, we can easily show that the limiting values of (30) and (31) are both equal to

$$\gamma^2(u_0) = \sigma^2\circ F^{-1}(u_0) \int_{-\infty}^{\infty} K(t)^2 dt \ .$$

Hence $(n\,h(n))^{1/2} B_n^*$ and $(n\,h(n))^{1/2}\hat{B}_n^*$, by Proposition 1 , are asymptotically equivalent in mean square. Now if we define

$$W_n^* = n^{-1} \sum_{i=1}^{n} Z_{in} ,$$

where

$$Z_{in} = (Y_i - m(X_i))h(n)^{-1}K\left(\frac{F(X_i)-u_0}{h(n)}\right) ,$$

then

$$(32) \qquad E[n\,h(n)\,(\hat{B}_n^*-W_n^*)^2]$$

$$= h(n)E\left\{(Y_1-m(X_1))^2[Q_n^*(F(X_1))-h(n)^{-1}K\left(\frac{F(X_1)-u_0}{h(n)}\right)]^2\right\}$$

Clearly the right hand side of (32) converges to zero and thus $(n\,h(n))^{1/2}\hat{B}_n^*$ and $(n\,h(n))^{1/2}W_n^*$ are asymptotically equivalent in mean square. W_n is sum of independent identically distributed random variables Z_{in} $(i = 1,\ldots,n)$ with zero means and $\lim_{n\to\infty} h(n)\mathrm{Var}(Z_{in}) = \gamma^2(u_0)$. It is also easy to check that the sufficient conditions for Liapounov's Central Limit Theorem for double sequence of random variables are satisfied for the Z_{in}. Therefore $(n\,h(n))^{1/2} W_n^*$ is asymptotically $N(0,\gamma^2(u_0))$, so is $(n\,h(n))^{1/2}\hat{B}_n$ and the theorem follows.

Theorem 6 can be used to construct approximate confidence interval for $\alpha oF^{-1}(u_0)$ if we can replace $E(M_n^*)$ by $\alpha oF^{-1}(u_0)$. In other words we need $\lim_{n\to\infty}(n\,h(n))^{1/2}[E(M_n^*) - \alpha oF^{-1}(u_0)] = 0$. The following corollary will allow us to do this when certain conditions are met.

Corollary 1. Suppose the assumptions of Theorem 6 are satisfied, and in addition the following conditions are also satisfied:

(a) $\int_{-\infty}^{\infty} zK(z)dz = 0$

(b) $\int_{-\infty}^{\infty} z^2 |K(z)| dz < \infty$.

(c) $K''(z)$ exists and satisfies (i) and (ii) of Theorem 5.

(d) The second derivative of $g(u) = \alpha o F^{-1}(u)$ exists and is continuous at u_0 .

Then $\lim\limits_{n\to\infty} [E(M_n^*)-g(u_0)]/h(n)^2 = g''(u_0) \int_{-\infty}^{\infty} K(z)dz$.

Furthermore, if $\lim\limits_{n\to\infty} h(n)n^{1/5} = 0$, then

$$\mathcal{L}[(n\,h(n))^{1/2}(M_n^*-g(u_0))] \to N(0,\gamma^2(u_0))$$

as $n \to \infty$, where $\gamma^2(u_0)$ is defined as in Theorem 6.

Proof. Write

$$|E(M_n^*) - g(u_0)|/h(n)^2$$

$$\leq |I_1| + |I_2| + |I_3| \quad , \text{ where}$$

$$I_1 = h(n)^{-2} \int_{u \notin [0,1]} h(n)^{-1} K\left(\frac{u-u_0}{h(n)}\right) g(u_0)\,du \;,$$

$$I_2 = h(n)^{-2} \int_0^1 h(n)^{-1} \sum_{i=1}^{n} \left[K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) - K\left(\frac{u-u_0}{h(n)}\right)\right] p(i;n,u) g(u)\,du \;,$$

$$I_3 = h(n)^{-2} \int_0^1 h(n)^{-1} K\left(\frac{u-u_0}{h(n)}\right) (g(u) - g(u_0))\,du \;,$$

and $p(i;n,u)$ is defined as in Lemma 1.

Clearly for sufficiently large n ,

$$|I_1| \le |g(u_0)| \int_{u \notin [-u_0/h(n),\,(1-u_0)/h(n)]} z^2 |K(z)|\,dz$$

and hence $|I_1|$ tends to zero as $n \to \infty$.

By the mean value theorem of the second order we may write

$$I_2 = h(n)^{-3} \int_0^1 \sum_{i=1}^{n} \left\{ \left(\frac{\frac{i}{n} - u}{h(n)}\right) K'\left(\frac{u-u_0}{h(n)}\right) + \frac{1}{2}\left(\frac{\frac{i}{n} - u}{h(n)}\right)^2 K''\left[\left(\frac{u-u_0}{h(n)}\right) + \theta_i \left(\frac{\frac{i}{n} - u}{h(n)}\right)\right]\right\}$$

$$p(i;n,u) g(u)\,du \;, \quad \text{where} \quad |\theta_i| \le 1 \;.$$

Since by assumption $K''(t)$ satisfies properties (i) and (ii) of $K(t)$, similar arguments used to derive (24) will enable us to show that

$$|I_2| \le \frac{1}{8(n\,h(n))^4} \left| \int_0^1 h(n)^{-1} K''\left(\frac{u-u_0}{h(n)}\right) g(u)\,du \right| \; .$$

$$+ \; 0(1/(n\,h(n))^4)$$

and this implies that $I_2 \to 0$ as $n \to \infty$.

Using the mean value theorem of the second order and assumption (a) , we may write

$$I_3 = g'(u_0) \int_{z \notin [-u_0/h(n),\,(1-u_0)/h(n)]} h(n)^{-1} zK(z)\,dz \quad +$$

$$\int_{-u_0/h(n)}^{(1-u_0)/h(n)} z^2 K(z) g''(\delta_n h(n) z + u_0)\,dz \; ,$$

where $|\delta_n| \le 1$. Since by assumption (b) , $\int_{-\infty}^{\infty} z^2 |K(z)|\,dz < \infty$, the first integral tends to zero as $n \to \infty$. By assumption, $g''(u)$ is continuous at u_0 . Hence the second integral tends to $g''(u_0) \int_{-\infty}^{\infty} z^2 K(z)\,dz$ as $n \to \infty$. Now clearly, the corollary follows.

<u>Remark 5.</u> Let

$$M^*_{n1} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) Y_{[i:n]} \; ,$$

$$M^*_{n2} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) I(Y_{[i:n]} \varepsilon A) \ ,$$

$$M^*_{n3} = n^{-1} \sum_{i=1}^{n} h(n)^{-1} K\left(\frac{\frac{i}{n} - u_0}{h(n)}\right) Y^2_{[i:n]} - (M^*_{n1})^2 \ .$$

If all the appropriate conditions are satisfied, then the approximate $(1-\alpha)100\%$ confidence intervals for $m \circ F^{-1}(u_0)$ and $P(Y \varepsilon A | X=F^{-1}(u_0))$ are respectively

$M^*_{n1} \pm [M^*_{n3} C/(nh(n))]^{1/2} \Phi^{-1}(\alpha/2)$ and $M_{n2} \pm [M^*_{n2}(1-M^*_{n2}) C/(nh(n))]^{1/2} \Phi^{-1}(\alpha/2)$

where $C = \int_{-\infty}^{\infty} K(t)^2 dt$ and Φ is the cdf of the standard normal random variables. An example of $K(t)$ fulfilled all the requirements of Corollary 1 is the pdf of the $N(0,1)$ variate. Note also that with this type of $K(t)$, Theorem 5 remains valid if we replaced condition (iv) by $\lim_{n\to\infty} h(n)^2 n = \infty$.

In order to have the same results for M_n, we need $E[n\, h(n)\, (M_n - M^*_n)^2] \to 0$ as $n \to \infty$. Inspection of the proof of $E(M_n - M^*_n)^2 \to 0$ as $n \to \infty$ in Theorem 5 leads immediately to the following corollary.

Corollary 2. Theorem 6 and Corollary 1 continue to hold for M_n if the condition (i) of Theorem 6 is replaced by the following condition:

There exists constants $M\ (>0)$ and $\alpha\ (>3/2)$ such that $|K(t_1) - K(t_2)| \leq M|t_1 - t_2|^\alpha$ for all t_1 and t_2.

Remark 6. An example of $K(t)$ possessed all the requirements of Corollary 2 is $t^2\exp(-t^2/2)/(2\pi)^{1/2}$. If all the appropriate conditions are satisfied, then the approximate $(1-\alpha)100\%$ confidence intervals for $m(x_0)$ and $P(Y\varepsilon A|X=x_0)$ are respectively $M_{n1} \pm [M_{n3}C/(nh(n))]^{1/2}\Phi^{-1}(\alpha/2)$ and $M_{n2} \pm [M_{n2}(1-M_{n2})c/(nh(n))]^{1/2}\Phi^{-1}(\alpha/2)$ where C and Φ are defined as in Remark 5.

3.2 Testing a specified regression function. The statistic R_n as defined in Remark 2 can be used to construct large sample tests concerning a specified regression function. For example we wish to test the null hypothesis $H_0 : m(x) = m_o(x)$ against the alternative hypothesis $H_1 : m(x) > m_o(x)$. By Remark 2 under H_0 $\mathcal{L}(n^{1/2}R_n/V) \to N(0,1)$ as $n \to \infty$, where $V^2 = \int_{-\infty}^{\infty} J(F(x))^2 \sigma^2(x)dF(x)$. Theorem 4 implies that

$$V_n = \sum_{i=1}^{n} n^{-1} J(\frac{i}{n+1})^2 [Y_{[i:n]} - m_o(X_{i:n})]^2$$

is a consistent estimator of V^2. Hence a large sample size α test is to reject H_0 if $n^{1/2}R_n/V_n \geq \Phi^{-1}(1-\alpha)$ where Φ is the standard normal c.d.f.

Footnotes

AMS 1970 subject classification. Primary 62E20, 62G30; Secondary 60F05, 62E15, 62G05, 62G10.

Key words and phrases. Concomitants of order statistics, asymptotic normality, consistent estimation, regression function.

References

Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. Ann. Statist. 2, 1034-1039.

Bhattacharya, P. K. (1976). An invariance principal in regression analysis. Ann. Statist. 4, 621-624.

Cacoullos, T. (1966). Estimation of a multivariate density. Ann. Inst. Statist. Math. 18, 178-189.

David, H. A. (1970). Order Statistics. Wiley, New York.

David, H. A. and Galambos, J. (1974). The asymptotic theory of concomitants of order statistics. J. Appl. Probability, 11, 762-770.

David, H. A., O'Connel, M. J. and Yang, S. S. (1977). Distribution and expected value of the rank of a concomitant of an order statistic. Ann. Statist. 5, 216-223.

Hajek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39, 325-346.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

Lehman, E. L. (1966). Some concepts of dependence. Ann. Math. Statist. 37, 1137-1153.

Mahalanobis, P. C. (1961). A method of fractile graphical analysis. Sankhyā Ser. A 23, 41-64.

Parthasarathy, K. R. and Bhattacharya, P. K. (1961). Some limit theorems in regression theory. Sankyā Ser. A. 23, 91-102.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

Sen, P. K. (1976). A note on invariance principles for induced order statistics. Ann. Probability 4, 474-479.

Stigler, S. M. (1969). Linear function of order statistics. Ann. Math. Statist. 40, 770-788.

Stigler, S. M. (1974). Linear function of order statistics. Ann. Statist. 2, 676-693.

Watson, G. S. (1964). Smooth regression analysis. Sankhyā Ser. A. 26, 359-372.

Yang, S. S. (1977). General distribution theory of the concomitants of order statistics. Ann. Statist. (to appear).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Linear Function of Concomitants of Order Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) S. S. Yang		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics M.I.T., Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-331)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE September 9, 1977
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Concomitants of order statistics, asymptotic normality, consistent estimation, regression function.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (see reverse side)		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Let (X_i, Y_i) $(i=1,2,\ldots,n)$ be i.i.d. as (X,Y). Then the r^{th} ordered X-variate is denoted by $X_{r:n}$ and the associated Y-variate, the concomitant of the r^{th} order statistic, by $Y_{[r:n]}$. This paper considers statistics of the form $n^{-1}\sum_{i=1}^{n} J(\frac{i}{n+1})Y_{[i:n]}$ and more generally of the form $n^{-1}\sum_{i=1}^{n} J(\frac{i}{n+1})H(X_{i:n}, Y_{[i:n]})$, where J is a bounded smooth function and may depend on n. Under certain regularity conditions, the asymptotic normality of these statistics are established. These statistics are used to construct consistent estimators of various conditional quantities, for example $E(Y|X=x)$, $P(Y \in A|X=x)$ and $Var(Y|X=x)$. Based on one of these statistics, a large sample test is also proposed for testing a specified regression function.

Y(r:n)

Sum over i from 1 to n

(i/(n+1))

an element of

bar